Enhancing Gender Classification by Combining 3D and 2D Face Modalities

Baiqiang Xia, Boulbaba Ben Amor, Huang Di, Daoudi Mohamed, Wang Yunhong, Drira Hassen

ENHANCING GENDER CLASSIFICATION BY COMBINING 3D AND 2D FACE MODALITIES

Baiqiang Xia[†], *Boulbaba Ben Amor*[‡], *Di Huang*[*], *Mohamed Daoudi*[‡], *Yunhong Wang*[*], *Hassen Drira*[‡]

[†] University Lille1, LIFL (UMR CNRS 8022), France.
[‡] Institut Mines-Telecom/TELECOM Lille1, LIFL (UMR CNRS 8022), France.
[*] Beihang University, IRIP Laboratory, China.

ABSTRACT

Shape and texture provide different modalities in face-based gender classification. Although extensive works have been reported in the literature, the majority of them are in the scope of shape or texture modality, individually. Among them, only a few concern their combination, and to the best of our knowledge, no work considers the combination with the 3D face surface. In this work, we investigate the combination of shape and texture modalities for gender classification, with both the combination of range images and gray images, and the combination of 3D meshes and gray images. In 10-fold subject-independent cross-validation with Random Forest on the FRGC-2.0 dataset, we achieved a correctness of $93.27\% \pm 5.16\%$, which outperforms each individual modality and is comparable to the state-of-the-art. Results confirm that shape and texture modalities are complementary, and their combination enhances the performance of face-based gender classification.

Index Terms— 3D/2D face modality, Gender classification, DSF, LBP, Random Forest, fusion.

1. INTRODUCTION

Human faces provide a variety of important demographic information, such as gender, age and ethnicity. The information is conveyed by different modalities, e.g., shape and texture, of face representation. Physically, the 3D face shape defines the solid border which distinguishes the face and the environment, while the texture represents the reflection and absorption effects of external illumination caused by the face skin. They are naturally attached in physical face and are used intuitively together in human face interpretation activities. Different modalities of human faces provide different cues for facial attribute recognition. In the literature of face-based gender recognition, a list of works have been done with consideration of these two modalities.

For texture-based gender classification, *Ylioinas et al.* [1] combine Contrast Information (strength of patterns) and Local Binary Patterns (LBP). In [2], *Yang et al.* fuse the outputs of the Active Appearance Models (AAM) and LBP with sequence selection algorithms. *Shan* selects discriminative LBP features with Adaboost in [3]. In [4], *Kumar et al.* classify gender with a set of visual attributes of face. Recently, in [5], *Wang et al.* enhance LBP with one of its variants, named Local Circular Patterns (LCP), for gender classification. In [6], *Makinen and Raisamo* make a comparative study of different gender classification methods, and find that the database, normalization, hair, and experiment settings account more for the results, than the classifiers. For shape-based gender classification, in [7], *Liu et al.* extract features in consideration of the height and orientation differences on symmetrical facial points. In [8], *Han et al.* extract geometric features with the volume and area information of faces. In [9], *Hu et al.* divide each face into four regions for feature extraction, and find that the upper face is the most discriminative in the face. In [10], *Toderici et al.* obtain features with the MDS (Multi Dimensional Scaling) and wavelets. In [11], *Ballihi et al.* select salient geometrical facial features with Adaboost. Recently, in [12], *Xia et al.* obtain Dense Scalar Field features through shape comparison of symmetrical faical curves.

Despite that the methods on individual texture or shape modality have acquired good performance in the literature, it has been noticed that the combination of shape and texture modalities tends to be a better strategy in face-based gender classification. In [13], *Lu et al.* use SVM (Support Vector Machine) to generate the posterior probabilities of range and intensity images. The probabilities are then fused with equal weight and compared directly for gender classification. In [14], *Wu et al.* combine shape and texture implicitly with needle maps recovered from intensity images in gender classification. In [15], *Huynh et al* fuse the Gradient-LBP features from range images, and the Uniform LBP features on 2D gray images, for gender classification. All these previous works have found that combination of shape and texture modalities yields better performance with their experimental results.

From the literature, we observe that the combination of shape and texture modalities for gender classification has been noticed, while as far as we know, no effort has been issued in combination of the 3D mesh and 2D texture image.

The previous studies in [13, 14, 15] are based on relative small datasets (less than 1250 3D images), leaving doubt on their statistical-significance on a more comprehensive dataset. In addition, experiments described in [13] and [14] are conducted on neutral faces. In contrast, our work investigates the relationship of shape and texture modalities not only with the combination of range and gray images, but also with the combination of 3D face meshes and gray images. Experiments are carried out on the FRGC-2.0 dataset, where as many as 4007 scans for each type of data (3D mesh, range image, gray image) exist. With these improvements, we ensure that our investigation on the combination of face shape and texture for gender classification is more thorough and statistical-significant. In our approach, we first perform gender classification for each type of data (mesh, range image and gray image), with Dense Scalar Field (DSF) features on 3D meshes, LBP features on range images and 2D gray images. Then both the combination of mesh and gray image, and combination of range and gray image, are explored in fusion scenario.

The rest of this paper is organized as follows. The details of DSF and LBP features for 3D and 2D are described in section 2 and 3, respectively. Random Forest classifier and the fusion schema are presented in section 4. In section 5, extensive experiments are shown to justify our approach. Finally, we draw conclusion in section 6.

2. EXTRACTING DSF FEATURES FROM 3D MESH

Regarding the anatomic founding in [16], where the authors find out that female faces usually possess smoother and rounder foreheads and cheeks, and less prominent noses, chins and jaws than male faces, we conclude that, generally, male faces possess more changes than female faces. In our work, as shown in Figure 1, we define a face template $\boldsymbol{T}$ as the middle point of the geodesic deformation path from a representative male face to a representative female face [17]. We can see that this face template $\boldsymbol{T}$ is more similar to the male face than to the female face. Thus, generally, more energy should be enforced to deform a female face to $\boldsymbol{T}$ than to deform a male face to $\boldsymbol{T}$, especially in the regions of nose, chins and jaws, which are more prominent in male faces.

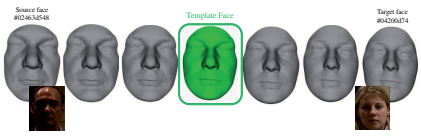

Fig. 1. The face template (the middle point of geodesic deformation path from a male scan to a female scan).

The energy needed to deform a face to $\boldsymbol{T}$ is represented by the *Dense Scalar Field (DSF)*. The methodology for extracting DSF features was first proposed by *Drira et al.* [18] for 3D face dynamic expression recognition, and then employed by *Xia et al.* [12] for extracting bilateral symmetric features in 3D face for gender classification. Formally, the face template $\boldsymbol{T}$ and the preprocessed 3D faces, which have undergone the same preprocessing procedures in [12] for hole-filling, cropping, smoothing and rotation, are first represented by a set of parameterized radial curves emanating from the nose tip. Then, for a pair of corresponding parameterized facial curves $\beta_{\boldsymbol{S}}^{\alpha}$ in a preprocessed face $\boldsymbol{S}$ and $\beta_{\boldsymbol{T}}^{\alpha}$ in the face template $\boldsymbol{T}$ with curve index α, where $\beta : I \rightarrow \mathbb{R}^3, I = [0, 1]$, they are represented as $q_{\boldsymbol{S}}^{\alpha}, q_{\boldsymbol{T}}^{\alpha} : I \rightarrow \mathbb{R}^3$, using the *square-root velocity function*, denoted by $q(t)$, where $q(t) = \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}$. Further, with the constraint of $\|q\| = 1$, where $\|\cdot\|$ implies the $\mathbb{L}^2$ norm, the length of $q_{\boldsymbol{S}}^{\alpha}$ and $q_{\boldsymbol{T}}^{\alpha}$ are unified. The tangent vector field $\dot{\psi}_{\alpha}^{*}$ on the geodesic path from $q_{\boldsymbol{S}}^{\alpha}$ to $q_{\boldsymbol{T}}^{\alpha}$ is given by (1):

$$\dot{\psi}_{\alpha}^{*} = \frac{\theta}{\sin(\theta)} \left(q_{\boldsymbol{T}}^{\alpha} - \cos(\theta) q_{\boldsymbol{S}}^{\alpha}\right) \tag{1}$$

where $\theta = cos^{-1}(\langle q_{\boldsymbol{S}}^{\alpha}, q_{\boldsymbol{T}}^{\alpha} \rangle)$, and $\langle ., . \rangle$ implies the $\mathbb{L}^2$ inner product of its elements. With $V_{\alpha}^{k} = ||\dot{\psi}_{\alpha}^{*}(k)||$, the magnitude of $\dot{\psi}_{\alpha}^{*}$ at each point, with curve index α and point index k, a *Dense Scalar Field* (DSF) between a preprocessed face $\boldsymbol{S}$ and the face template $\boldsymbol{T}$ is extracted. This Dense Scalar Field quantifies the shape difference between corresponding radial curves in $\boldsymbol{S}$ and $\boldsymbol{T}$ on each indexed point. Figure 2 examplifies this DSF. For each subject, the face in column (a) shows the 2D intensity image; column (b) illustrates the preprocessed 3D face surface $\boldsymbol{S}$; column (c) shows the 3D face $\boldsymbol{S}$ with extracted curves; column (d) shows color-map of the DSF mapped on $\boldsymbol{S}$ with three poses. In echoing with what we stated in the beginning of this section, we observe that, female faces generally need more energy to deform to the face template $\boldsymbol{T}$, especially in the nose, chin and jaw regions.

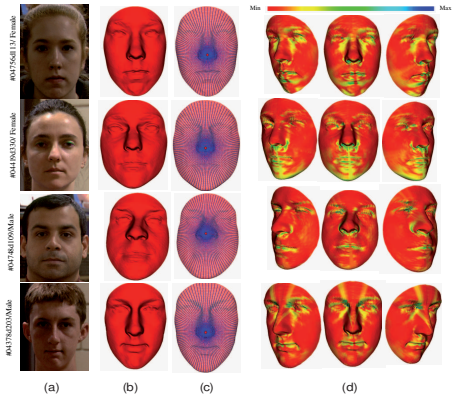

Fig. 2. Illustrations of DSF on face. (a) 2D intensity image; (b) preprocessed 3D face surface $\boldsymbol{S}$; (c) $\boldsymbol{S}$ with extracted curves; (d) color-map of the DSF mapped on $\boldsymbol{S}$ with three poses.

3. EXTRACTING LBP FEATURES FROM RANGE AND GRAY IMAGES

LBP, a non-parametric method [19], was originally proposed to describe local texture variations in 2D images. The basic LBP operator labels image pixels by thresholding in a 3×3 neighborhood. If the value of a neighbor pixel is no lower than the central pixel, the corresponding binary bit is set to 1; otherwise to 0. By concatenating all the eight binary bits, a binary number is hence formed, and the corresponding decimal number is used for labeling. Due to its discriminative power, tolerance to monotonic lighting changes and computational simplicity, LBP has been extensively adopted in 2D face recognition [20]. Some recent works also reveal its competence for face-based gender classification [5, 21].

As noted in [22], when the LBP operator works on texture, the LBP codes are regarded as micro-textons such as curved edges, spots, flat regions, etc. Similarly, when LBP operates on range images which are based on depth information, it describes local shape structures, e.g. flat, concave, convex etc. In our work, we use the LBP operator to extract the gender related cues on both range and gray images, for the purpose of comparison and possible combination. Technically, first, facial images (range or gray images) are cropped and resized to a unified size in preprocessing. Then with the preprocessed image divided into several local regions, the LBP bins of each region are extracted on each pixel within a neighborhood of 3×3 pixels. Finally, the LBP bins on each region are categorized into 59 Uniform LBP patterns, and the histograms of these Uniform LBP patterns on each region are generated and concatenated, to form a face description which contains both local and global information. Figure 3 shows some examples of LBP histogram on range and gray images.

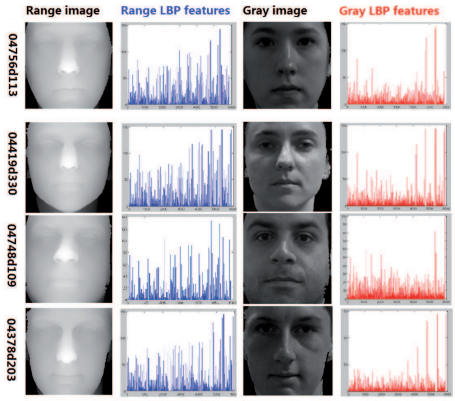

Fig. 3. Illustrations of LBP features. For each subject, from left to right, the columns are: the range image, the concatenated LBP histograms of the range image, the gray image, and the concatenated LBP histograms of the gray image.

4. GENDER CLASSIFICATION

Face-based gender classification is a binary problem which classifies the gender of query face into male or female. We carry out gender classification experiments with the well-known machine learning algorithm, Random Forest. Random Forest is an ensemble learning method that grows many classification trees. To classify a new subject represented by an input feature vector, each tree of the forest gives a classification result and the forest makes the predicted class label by choosing the classification having the most votes. The forest also generates a probability for the predicted class label.

With the predicted class labels and probabilities given by Random Forest, we define a fusion strategy named maximum-criteria as following: suppose that for a query subject, the forest generated with one type of features makes a predicted class label *A* with a probability *Pa*, and the forest created with another type of features yields a predicted class label *B* with a probability *Pa*. If $Pa \geq Pb$, the predicted class label of the fusion comes to *A*; otherwise the fusion result is *B*.

5. EXPERIMENTS

The FRGC-2.0 database was collected by researchers from the University of Notre Dame. It contains 4007 3D face meshes constructed from 4007 range images, and 4007 corresponding 2D scans of 466 subjects with differences in gender, ethnicity, age and expression [23]. For gender, there are 1848 scans of 203 female subjects and 2159 scans of 263 male subjects. With FRGC-2.0, we performed a set of 10-fold subject-independent cross-validation experiments with 100-tree random forest. First, we performed gender classification experiments considering only one modality, with LBP features from gray images, LBP features from range images, and DSF features from 3D meshes, respectively. Then, using the maximum-criteria, we experimented the combination of shape and texture modalities in two ways: one fuses the results from gray images and range images, and the other fuses the results from gray images and 3D meshes.

Experimental results are shown in Figure 4. The bars and the vertical lines illustrate the recognition rates and the standard deviations. From Figure 4, we observe that results with all single modalities are relatively effective ($> 85\%$). The DSF features extracted from 3D mesh perform better than the LBP features extracted from 3D range image ($90.50\% > 88.13\%$), and LBP features yeild better performance with 3D range images than with 2D gray images ($88.51\% > 86.47\%$). In fusion scenario, both the combination of 3D-mesh/gray-image, and the combination of range-image/gray-image, outperfom its corresponding individual modality. It confirms that shape and texture modalities are complementary in face-based gender classification. The best result is achieved in combination with 3D mesh and 2D gray image, with a mean correctness of 93.27% and a standard deviation of 5.16%.

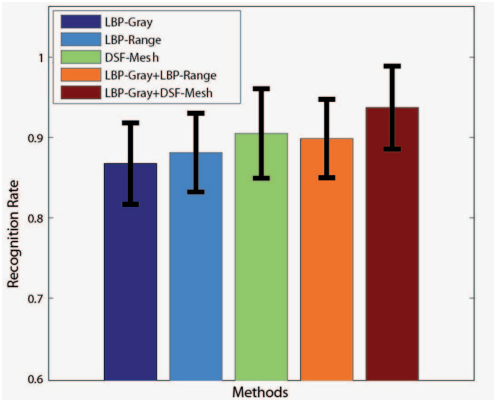

Fig. 4. Gender classification results. The bars signify the recognition rates and the vertical lines signify the corresponding standard deviations in experiments. From left to right, the columns are: the result with LBP features from gray image, the result with LBP features from range image, the result with DSF features on 3D meshes, the result with combination of LBP features from gray and range image, and the result with combination of LBP features from gray images and DSF features from 3D meshes.

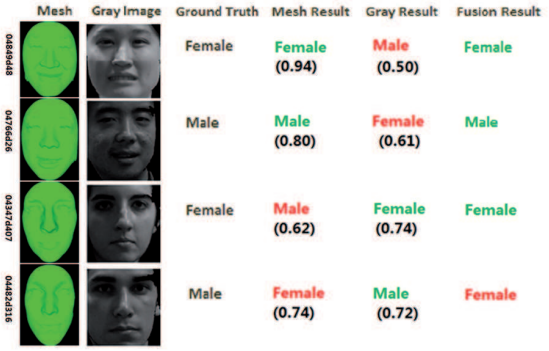

Fig. 5. Fusion with results from meshes and gray images. For each subject, from left to right, the columns are: the preprocessed 3D mesh, the preprocessed gray image, the ground truth of gender, the predicted class label and the probability with 3D mesh, the predicted class label and the probability with gray image, the predicted class label of fusion.

Figure 5 and Figure 6 present some examples in fusion scenario. From left to right, the columns are: the preprocessed 3D data (mesh for Figure 5 and range image for Figure 6), the corresponding preprocessed 2D gray image, the ground truth of gender, the predicted class label and the probability with 3D data (meshes for Figure 5 and range images for Figure 6, with both predicted class label and its probability), the predicted class label and the probability with gray image (predicted class label and its probability), and the predicted class

Fig. 6. Fusion with results from range and gray images. For each subject, from left to right, the columns are: the preprocessed range image, the preprocessed gray image, the ground truth of gender, the predicted class label and the probability with range image, the predicted class label and the probability with gray image, the predicted class label of fusion.

label of fusion. We observe that in both Figure 5 and Figure 6, there are cases where the classification one of the modalities fails, but the fusion makes the correct prediction. These cases explain directly why improvements occur in fusion.

Table 1. comparison with state of the art.

Author	*This work*	*Lu et al. [13]*	*Huynh et al. [14]*	*Wu et al. [15]*
Data	*mesh+gray*	*range+gray*	*range+gray*	*needle map*
Data Size	*4007 scans/ 466 subjects*	*1240 scans/ 376 subjects*	*1149 scans/ 105 subjects*	*260 scans*
Result	$93.27 \pm 5\%$	$91 \pm 3\%$	96.70%	$93.6 \pm 4\%$

Table 1 gives a comparison to the highly related works. Although we have experimented with significantly more scans and more subjects, our recognition rate is higher than the work [13], and comparable to the work [15]. The recognition rate is lower than the work of *Huynh et al.* [14], while it is should be noted that their result is based on 1149 pairs of range and gray images of 105 subjects, far less than in our work, where 4007 pairs of 3D meshes and gray images of 466 subjects are involved.

6. CONCLUSION

In this paper, we have investigated the combination of face shape and texture modalities for gender classification. With 3D DSF features calculated directly from 3D meshes, and the LBP features extracted from range and gray images, the fusions of 3D-mesh/gray-image and range-image/gray-image have always demonstrated the superiority to their correspond-

ing individual modalities on the FRGC-2.0 dataset. We have achieved a comparable result of $93.27 \pm 5.16\%$, in combination of 3D meshes and gray images. Results confirm that shape and texture modalities are complementary, and the combination of them performs better than individual modality, in face-based gender classification.

7. REFERENCES

[1] J. Ylioinas, A. Hadid, and M. Pietikinen, "Combining contrast information and local binary patterns for gender classification," in *Image Analysis*, 2011, vol. 6688, pp. 676–686.

[2] W. Yang, C. Chen, K. Ricanek, and C. Sun, "Gender classification via global-local features fusion," in *Biometric Recognition*, 2011, vol. 7098, pp. 214–220.

[3] C. Shan, "Learning local binary patterns for gender classification on real-world face images," in *Pattern Recognition Letters*, 2012, vol. 33, pp. 431–437.

[4] N. Kumar, A. Berg, P.N. Belhumeur, and S. Nayar, "Describable visual attributes for face verification and image search," in *Pattern Analysis and Machine Intelligence*, 2008, vol. 33, pp. 1962 –1977.

[5] C. Wang, D. Huang, Y. Wang, and G. Zhang, "Facial image-based gender classification using local circular patterns," in *21st International Conference on Pattern Recognition*, 11 2012.

[6] E. Makinen and R. Raisamo, "An experimental comparison of gender classification methods," in *Pattern Recognition Letters*, 2008, vol. 29, pp. 1544–1556.

[7] Y. Liu and J. Palmer, "A quantified study of facial asymmetry in 3D faces," in *Analysis and Modeling of Faces and Gestures*, 2003, pp. 222–229.

[8] X. Han, H. Ugail, and I. Palmer, "Gender classification based on 3D face geometry features using svm," in *CyberWorlds*, 2009, pp. 114–118.

[9] Y. Hu, J. Yan, and P. Shi, "A fusion-based method for 3D facial gender classification," in *Computer and Automation Engineering (ICCAE)*, 2010, vol. 5, pp. 369–372.

[10] G. Toderici, S. O'Malley, G. Passalis, T. Theoharis, and I. Kakadiaris, "Ethnicity- and gender-based subject retrieval using 3-D face-recognition techniques," in *International Journal of Computer Vision*, 2010, vol. 89, pp. 382–391.

[11] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3D-geometric features for efficient face recognition and gender classification," in *IEEE Transactions on Information Forensics & Security*, 2012, vol. 7, pp. 1766–1779.

[12] B. Xia, B. Ben Amor, H. Drira, M. Daoudi, and L. Ballihi, "Gender and 3D facial symmetry: What's the relationship?," in *IEEE Conference on Automatic Face and Gesture Recognition*, 2013.

[13] X. Lu, H. Chen, and A. Jain, "Multimodal facial gender and ethnicity identification," in *Proceedings of the 2006 international conference on Advances in Biometrics*, 2006, pp. 554–561.

[14] J. Wu, W.A.P. Smith, and E.R. Hancock, "Gender classification using shape from shading," in *International Conference on Image Analysis and Recognition*, 2007, pp. 499–508.

[15] T. Huynh, R. Min, and J.L. Dugelay, "An efficient lbp-based descriptor for facial depth images applied to gender recognition using rgb-d face data," in *ACCV 2012, Workshop on Computer Vision with Local Binary Pattern Variants*, 2012.

[16] "The main differences between male and female faces?," in *url:www.virtualffs.co.uk*.

[17] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3d face recognition under expressions, occlusions and pose variations," in *Pattern Analysis and Machine Intelligence*, 2013, vol. 99.

[18] H. Drira, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, "3D dynamic expression recognition based on a novel deformation vector field and random forest," in *21st International Conference on Pattern Recognition*, 2012.

[19] T. Ojala, M. Pietikinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in *Pattern Analysis and Machine Intelligence*, 11 2002, vol. 24, p. 971987.

[20] T. Ahonen, A. Hadid, and M. Pietikinen, "Face recognition with local binary patterns," in *ECCV*, 2004.

[21] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *ICB*, 2007.

[22] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: a survey," in *IEEE Trans. on Systems, Man, and Cybernetics*, 2011, vol. 41, pp. 765–781.

[23] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 947 – 954.